

The light-harvesting chlorophyll *a*–*c*-binding protein of dinoflagellates: a putative polyprotein

Roger G. Hiller*, Pamela M. Wrench, Frank P. Sharples

School of Biological Sciences, Macquarie University, Macquarie, NSW 2109, Australia

Received 6 February 1995; revised version received 8 March 1995

Abstract The principle light-harvesting chlorophyll *a*–*c*-binding protein of *Amphidinium carterae* of 19 kDa is encoded as a polyprotein translated from a 6.1 kb mRNA. The cDNA sequences indicate that each derived polypeptide is contiguous with the next and that the mature peptides are formed by cleavage at a C-terminal arginine residue. Comparison of the amino-acid sequences shows the *Amphidinium* protein to be most closely related to the fucoxanthin-chlorophyll-protein (Fcp) of *Phaeodactylum* and less related to the chlorophyll *a*–*b*-binding (Cab) proteins including those from *Euglena*.

Key words: Amino acid sequence; Chlorophyll *c*; Polyprotein; Light-harvesting complex; Dinophyceae; *Amphidinium carterae*

1. Introduction

The intrinsic light-harvesting proteins of higher plants and green algae binding Chl *a* and *b* are encoded by multigene nuclear families [1–3]. Many of the genes are expressed although the specific roles of many of the proteins are by no means clear. In the chromophyte algae the light-harvesting proteins bind Chl *a* and *c* and are also encoded by multigene families [4,5]. All these light-harvesting proteins share conserved motifs although the level of identity is overall rather low. It has, however, been suggested that they share a three transmembrane helix structure. This may indicate a common evolutionary origin involving duplication and fusion of genes encoding a two transmembrane protein followed by elimination of one of the four helices [4].

Peptides comprising the main light-harvesting complex of dinoflagellates have been previously investigated [6,7]. The principle polypeptide of 19 kDa is related at the N-terminal region to the light-harvesting proteins of other chromophytes as well as to the Cab proteins [6]. During PCR experiments designed to confirm synthesis of a cDNA extending to the known N-terminus of the mature 19 kDa peptide, an amplified band of ~1 kb was obtained in addition to that expected at 420 bp. The 1 kb band was found to encode a complete amino acid sequence of a 19 kDa polypeptide flanked by partial sequences encoding two other forms of the 19 kDa polypeptide.

We report here evidence, based on cDNA sequencing and Northern blotting, that a gene for the light-harvesting complex

of dinoflagellates encodes a polyprotein which is posttranslationally cleaved to up to 10 different 19 kDa mature polypeptides.

2. Materials and methods

Amphidinium carterae CS-21 was cultured, its main light-harvesting complex isolated and the 19 kDa principle polypeptide sequenced as previously described [6]. Total RNA was extracted by homogenising exponentially growing cells in Total RNA Isolation Reagent (Advanced Biotechnologies) using 1 ml reagent for 25 μ l of packed algal cells. After extraction of the lysed cells with chloroform, RNA was precipitated with isopropanol and washed with 75% ethanol according to the manufacturers' protocol. Precipitated RNA was redissolved in DEPC-treated water and used immediately for poly(A)⁺ mRNA preparation or precipitated by addition of 0.1 vol. of 3 M sodium acetate and 2.5 vol. of ethanol and stored at –20°C until required for Northern blotting.

Poly(A)⁺ mRNA was prepared from total RNA by addition of 0.1 vol. of Oligotex poly-T affinity reagent (Quiagen Inc.) following the manufacturers' protocol. First strand cDNA synthesis was by reverse transcription using the primer 2360 5'-AATTCGCGGCCGC(T)15 and a 5'-Amplifinder RACE kit (Clontech Inc). Fragments of cDNA were amplified by PCR using primers designed from protein sequence data for the 19 kDa principle peptide of the LHC. The forward primer for 1 kb product was 218 5'-TTCGACCCCTCGGTTTCAC corresponding to the amino acid sequence FDLGFT and the reverse primer 682 5'-CAT(ATGC)CC(TGA)AT(TGA)AT(ATGC)GCCATCAT corresponding to the amino acid sequence MMAIHGM. The 3'-end of the cDNA was amplified using as forward primer 2361 5'-CGGTGACTGGGCAACTACA designed to the 3'-end of the 1 kb cDNA product with the reverse primer 2360. Standard PCR conditions were 2 mM Mg, 100 μ M dNTP and 2 units each of Taq polymerase and Taq extender (Stratagene) in a total volume of 50 μ l. After an initial cycle of 3 min at 94°C, 1 min at 52°C and 2 min at 72°C the PCR reactions were given 34 cycles for 1 min, 1 min and 2 min at the above temperatures. Amplified PCR products were separated on 0.8% agarose gels, excised and purified with GeneClean II (Bio 101) and cloned into the pGEM-T vector (Promega) following the manufacturers protocols. DNA sequencing was by the dideoxy chain termination technique [9] using a Sequenase II kit (USB) and the manufacturer's protocol for double-stranded templates. Both strands were sequenced by a combination of standard subcloning and custom oligonucleotides. DNA and protein sequences were analysed using the Wisconsin program package accessed through the Australian National Genomic Information Service.

For Northern blotting RNA was dissolved in 25 mM EDTA, 0.1% SDS, size-fractionated on a 1.2% agarose gel and probed [8]. The probe for LHC was from the cloned 1 kb PCR fragment (pRGH200) cut with *Sac*I and contained 756 bp of the insert. For PCP the probe was a 1272 bp fragment obtained by genomic PCR encoding 131 C-terminal residues, the intergenic region and the N-terminal leader sequence plus 72 residues of an adjacent PCP (Hiller, Wrench and Sharples, unpublished observations). Probes were labelled using a Gigaprime kit from Biotec.

3. Results

The sequence of the cloned (pRGH200) original 1 kb amplified cDNA band together with a derived amino acid sequence

*Corresponding author. Fax: (61) (2) 850-8245.

Abbreviations: Chl, chlorophyll; LHC, light-harvesting complex; PCP, peridinin chlorophyll *a*-protein; Cab, chlorophyll *a*–*b*-binding; PCR, polymerase chain reaction; Fcp, fucoxanthin-chlorophyll-protein.

The sequences reported here will appear in the EMBL nucleotide data base under numbers Z47562 and Z47563.

	S				Q				
a	XFENERGVQD	PVGFFDPLGF	TADGSVENFK	RLAQTEIKHG	RVAMLATMGY				50
b	a--l--a	-t-----	ss--d--	-rras--	-----				
c	a--s-l--aa	-t-w----l	ak--mka--	-rras--	-i-----				
d		-----	ke-n-d--	-rra-----	-i-----				
	S								
a	ITPEITGKLP	GYLSPSTGVK	YDDIPNGLGA	ISKVPAAGWG	QMIAYAAFSE				100
b	m-----a-f	-----y-qss-	fa-v-----	msk-----vl--a	-va--g-vcd				
c	---l---f-	---m-l-	-e-----	---fv--a	-il---fyc-				
d	---l#frfd	---r-a--	-a-----a-	f---fv--v	--v--c1lv-				
	I				G		A		
	S		A		A		PI		
							P		
a	LSQDQSAGTP	AAEGDFGFKV	LTSSDPAELE	KKLSAELANG	RLAMMAIIGM				150
b	v-----g-a	-----	i--e-eet-k	r--ns-----	-----l				
c	q-----e-sa	gea-----	---k-eeg-	r--ns-i---	-----				
d	i-----q-a-	-q#-----	-----k-	---r-----	-----				
	FFQDGLTGSA		WGDWASYTAS		PL				
a	-----g-		y-----n--d-		--r				
b	-----y-		n-----f-		--r*				
c	-----		n-----e-		--r				

Fig. 3. Amino acid sequences of principle polypeptides of *Amphidinium* LHC. a, protein-derived sequence with major variations in uppercase letters above; b, complete cDNA-derived sequence; c, complete cDNA-derived sequence (3'-end of gene); d, partial cDNA-derived sequence. The masses of peptides b and c are 18,420 and 18,604, respectively. (-, identity; #, a space; *, stop codon).

elutes in the region of PTH-Met. We hypothesise, based on the cDNA derived sequences that X is a modified Ala. (It coelutes with PTH-*n*-methyl-alanine.) The Ser/Asn ambiguity at residue 4 results from approximately equal amounts of the two species and both are specified in the derived sequences. For many of the other variants we have not yet found correspondence in the derived sequences including 3 additional partial DNA sequences obtained by genomic PCR (Hiller and Wrench, unpub-

lished observations). The identity between the protein sequence and the cDNA-derived sequences is 71–78%, which is similar to that between the cDNA-derived sequences themselves.

4. Discussion

As was previously noted [6] comparison of the *Amphidinium* LHC sequence with those of other LHCs (Fig. 4) indicates

```

100
a .....
p .....
e .....
u .....
t .....
m .....

51
a .....
p .....
e .....
u .....
t .....
m .....

100
a .....
p .....
e .....
u .....
t .....
m .....

150
a .....
p .....
e .....
u .....
t .....
m .....

200
a .....
p .....
e .....
u .....
t .....
m .....

250
a .....
p .....
e .....
u .....
t .....
m .....

300
a .....
p .....
e .....
u .....
t .....
m .....

```

Fig. 4. Comparison of Cab and Chl *a-c*-binding protein sequences aligned using the PILEUP program. a, *Amphidinium* LHC (this paper); p, *Phaeodactylum* Fcp [5]; e and u, *Euglena* LHCPII [9]; t, Tomato LHCPI [11]; m, *Mantoniella* LHCP [12]. (*, identity; ·, a space.)

significant identity especially towards the N-terminus. The greatest identity is with *Phaeodactylum* Fcp (39%) whereas it is less than 30% with those of the Cab proteins. The atypical LHC of *Mantoniella* which binds Chl *c* as well as Chl *a* and *b* is only 22% identical and is clearly not part of the chromophyte evolutionary line as previously noted [13]. In a preliminary analysis of the *Amphidinium* sequence [6] we failed to find evidence based on the FDPLG motif for duplication of the sequences which include the postulated first and third transmembrane helices. Such a duplication, although not as strong as for the Cab proteins, is now apparent as shown below:

SVENFKRLAQTEIKHGRVAMLATMGYITPEIT

: : : : : :

AELEKKLSA--ELANGRLAMMAIIGMFFQDGL

A feature of the *Amphidinium* amino acid sequences is a high variability in Arg and Lys residues and the resulting multiplicity of peptides following tryptic digestion accounts in part for the shortcomings of our previous analysis.

There is only one comparable case (*Euglena* [10]) of an LHC being synthesised as a polypeptide and posttranslationally cleaved to multiple forms of the mature peptide. Taking *Euglena* LHCPII as a model [10], *Amphidinium* LHC mRNA of 6.1 kb would be made up of approximately 1 kb of the leader sequence, non-coding 5' and 3' regions plus the poly(A)⁺ tail. The remaining 5.1 kb could comprise up to ten 516/519 bp pseudo-repeats which encode the 172/173 amino acid LHC polypeptides. A major difference exists between the *Amphidinium* polypeptide, which is cleaved at an Arg residue without liberation of spacer peptide and the *Euglena* polypeptide. In *Euglena* the processing protease(s) recognises the consensus sequence XMXAXGXKX and liberates the corresponding decameric peptide. A feature of the gene encoding *Euglena* LHCPII is the occurrence of numerous introns which are flanked by atypical base sequences [10]. If the *Amphidinium* LHC gene contains introns they are likely to be fewer than in *Euglena* as Fig. 2 did not show any hybridisation to distinct species greater than 6.1 kb even after extended exposure. However, it is worth noting that the PCR artefact, which gave our original 1 kb PCR fragment (where the forward 218 oligonucleotide primer also functioned in reverse), has also been observed when the 218/682 primer combination was used for genomic PCR although in this case the site was quite different (Hiller and Wrench, unpublished observations). The last two 3' bases of primer 218 are AC which would hybridise to the GT 5' consensus of typical eukaryotic splice junctions.

Despite the low identity of *Amphidinium* LHC with LHCs containing Chl *b*, alignment of the sequence with the 3-dimensional structure of Kühlbrandt et al. [14] gives some insight into the binding of pigments. Within the sequences comprising the

first and third (B and A in the Kühlbrandt model) helices there are five sites binding Chl *a* which are identical. These are the Arg/Glu combinations which bind Chl *a*₄ and Chl *a*₁, the His binding Chl *a*₅, the Asn binding Chl *a*₂ and the Gln binding Chl *a*₃. It may be expected, due to the highly conserved nature of the inner antennae of the photosystems, that the position of some Chl *a* molecules, especially those at the lowest energy levels, will be common to LHCs in all organisms if the three transmembrane model is universally applicable. There may also be non-conserved Chl *a* binding sites if specific LHC peptides can only participate in energy transfer within the LHC.

Most phylogenetic dendrograms have euglenoids well separated from dinoflagellates [15,16] despite both having three membranes surrounding the chloroplast and we suggest that the unusual transport and processing of their respective LHCs has arisen independently. However, dinoflagellates may well be polyphyletic with respect to their plastids [16] and *Amphidinium carterae* may not be representative of all dinoflagellates.

Acknowledgements: We thank Dr. J. Speirs for assistance with Northern blotting, Dr. A.A. Gooley of Macquarie University Centre for Analytical Biotechnology for protein sequencing and Prof. Tony Larkum for commenting on the manuscript. The work was supported by an Australian Research Council Grant (A19131108) to R.G.H.

References

- [1] Thornber, J.P., Morishige, D.T., Anandan, S. and Peter, G.F. (1991) in: *Chlorophylls* (Scheer, H. ed.) pp. 549–585, CRC Press, Boca Raton, FL, USA.
- [2] White, M.J., Fristensky, B.W., Falconet, D., Childs, L.C., Watson, J.C., Alexander, C., Roe, B.A. and Thompson, W.F. (1992) *Planta* 188, 190–198.
- [3] Di Paolo, M.L., Dal Belin-Peruffo, A. and Bassi, R. (1990) *Planta* 181, 275–286.
- [4] Green, B.R. and Pichersky, E. (1994) *Photosynthesis Res.* 39, 149–162.
- [5] Grossman, A.R., Manodori, A. and Snyder, D. (1990) *Mol. Gen. Genet.* 224, 91–100.
- [6] Hiller, R.G., Wrench, P.M., Gooley, A.A., Shoebridge, G. and Breton, J. (1993) *Photochem. Photobiol.* 57, 125–131.
- [7] Inglesias-Prieto, R., Govind, N.S. and Trench, R.K. (1993) *Phil. Trans. R. Soc. Lond. B* 340, 381–392.
- [8] Fourney, R.M., Miyakoshi, J., Day III, R.S. and Paterson, M.C. (1988) *Focus* 10, 5–7.
- [9] Sanger, F., Nichlen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- [10] Muchhal, U.S. and Schwartzbach, S.D. (1992) *Plant Mol. Biol.* 18, 287–299.
- [11] Norris, B.J. and Miller, D.J. (1994) *Plant Mol. Biol.* 24, 673–677.
- [12] Pichersky, E. and Green, B.R. (1990) in: *Progress in Photosynthesis Research*, vol III (Baltscheffsky, M. ed.) pp. 553–556, Kluwer, Dordrecht.
- [13] Rheil, E. and Morschel, E. (1993) *Mol. Gen. Genet.* 240, 403–413.
- [14] Kühlbrandt, W., Wang, D.N. and Fujiyoshi, Y. (1994) *Nature* 367, 614–621.
- [15] Cavalier-Smith, T. (1989) *Nature* 339, 100–101.
- [16] Martin, W., Somerville, C.C. and Loiseaux-de Goer, S. (1992) *J. Mol. Evol.* 35, 385–404.